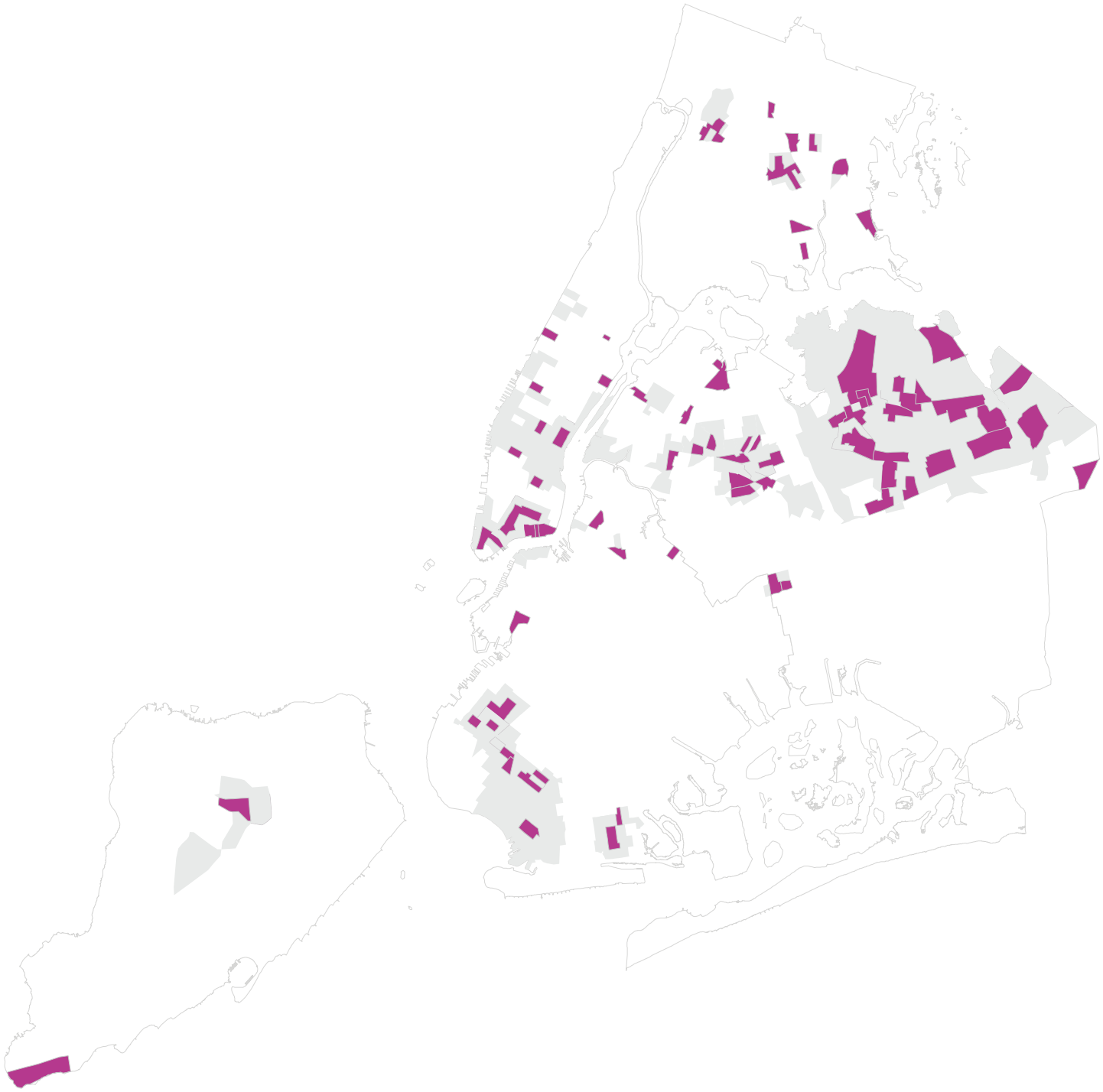
Creating Asian American Clusterhoods

Jean Kim and Ri Le
Geographic Information Systems • Leah Meisterlin • Fall 2018
Columbia GSAPP



Clusterhoods, all ethnicities

Contents

The model minority monolith myth, and social inequality amongst Asian Americans

We begin with a brief discussion of the origins of our research endeavor. Our project arose from an interest in illuminating the social realities of immigrants and people of color, as well as a shared interest in using GIS to understand our own social contexts and collective histories as Asian Americans, children of immigrants, and as residents and witnesses of New York City.

We recognize race as a social construct created to justify the unequal distribution of power and resources (Omi and Winant 2014). Through data collection methods, the popular discourse about Asian Americans as a monolithic group of "model minorities" is simultaneously imagined and constructed into a social reality. These methods that reduce all Asian ethnic groups into the "Asian or Pacific Islander" category racializes and quantifies a diverse group of peoples into a reductive category that is idealized and used as the middle-upper-class, hyper-educated, East Asian antithesis to other Black and Latinx groups who are imagined as deficient and underachieving.

However, this monolithic construction—a phenomena coexistent with the notion of Asian people being the "model minority"—ignores the inequality that divides ethnic groups lumped into the "Asian" category. For example, the White House AAPI Initiative in 2012 acknowledged that recently disaggregated Census data shows "staggering educational needs" of certain Asian ethnic groups that have historically been hidden by aggregated data collection methods. In terms of educational indicators, the report notes that "29 percent of Vietnamese-Americans, 38 percent of Hmong-Americans, 33 percent of Laotian-Americans, and 37 percent of Cambodian-Americans do not complete high school" (Aina 2012). In another example, a report from the Pew Research Center shows the growing inequality gap between the top 10% of Asians and the rest (known as the 90/10 ratio), with this income gap being the highest compared to all other racial groups (Kochhar, Rakesh, and Cilluffo 2018). The report cites immigration patterns and policies that dictated the influx of immigrants from various geopolitical circumstances as one of the causes of these inequalities.

Advances in research: the movement to disaggregate

In recent decades, scholars and activists have attempted to counter this monolithic imaginary of Asian American identity, calling for the disaggregation of the racial category of "Asian" in various realms including education and public health.

In the realm of higher education management and policy, the Count Me In campaign advocated by students of the University of California (UC) system pressured the institution to collect disaggregated admissions data on students and to create a separate Pacific Islander category in order to acknowledge the vast educational disenfranchisement of Pacific Islanders when compared to other "Asian" ethnic groups. Disaggregated data would allow the UC system to provide financial support and outreach projects that specifically targets these disadvantaged communities (Dizon 2011). Although disaggregation by ethnic groups is the most common mode, Dizon notes that other modes of disaggregation (such as languages spoken at home, immigration histories, and access to economic and social capital) reveals additional inequalities and differences between groups under the AAPI umbrella (Dizon 2011).

The Community Health Resource and Needs Assessment (CHRNA), developed by the NYU Center for the Study of Asian American Health, is an example of how this data, when disaggregated through other factors, can have major implications

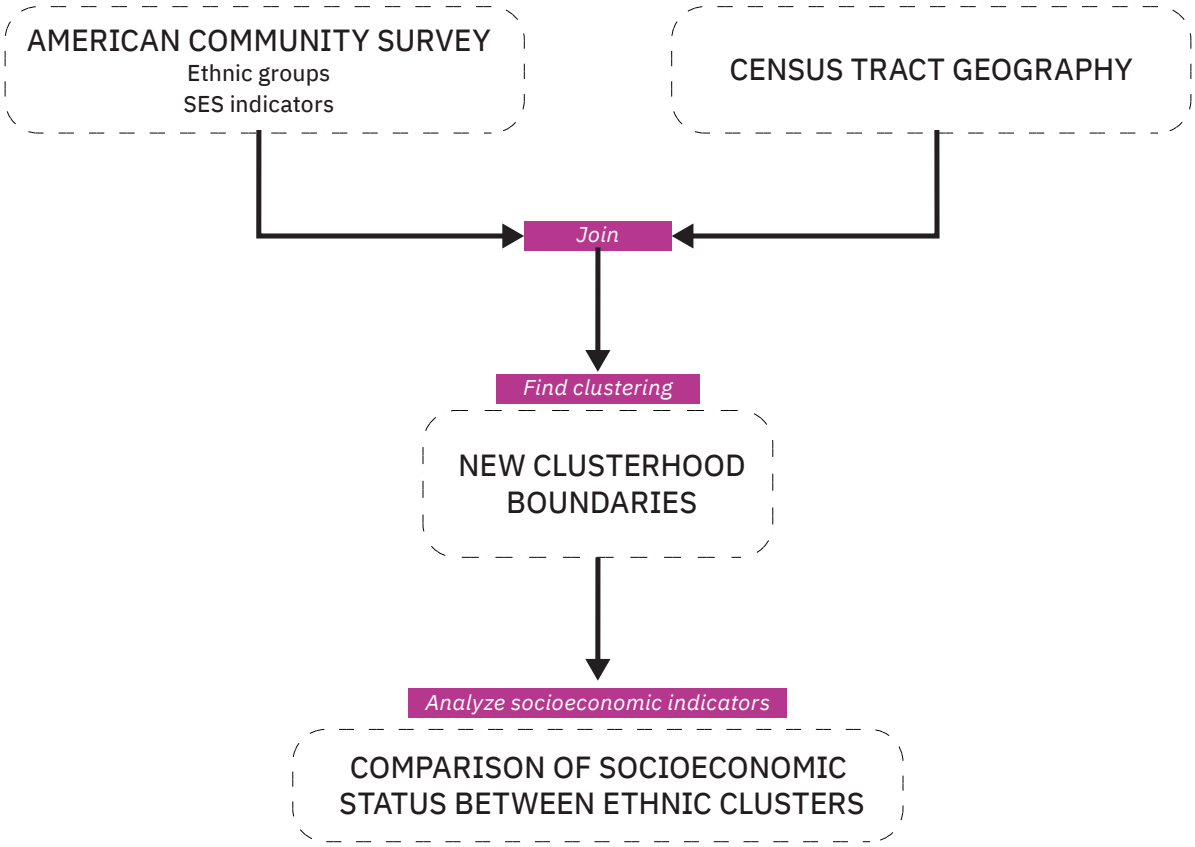
for public health policy. In order to collect the data, researchers collaborated with local community groups at specific sites with cultural or social significance. The identification of distinct ethnic groups enabled the researchers to segment their research and deliver reports that identified the unique needs of specific communities. For example, researchers identified mental health as the primary concern for Cambodian respondents, and specified the need for mental health services (that were culturally and linguistically sensitive) for Cambodian women ages 45 and up (NYU Center for the Study of Asian American Health 2016).

This movement to disaggregate data is reflected on the federal level, with a pending Senate Assembly Bill A7352 which proposes requirements in New York State for data collection, which explicitly includes languages spoken, country of birth, ancestry by ethnic group, and the collection of data on specific Hawaiian and Pacific Islander groups, citing the "diverse social, educational, health, and economic differences that are unique to their respective communities" as the rationale (NY State Senate, 2017).

Although exemplary, these examples are exceptions to the growing body of work on Asian Americans. Our project seeks to explicitly explore considerations of specific GIS techniques as they may be applied to research on Asian Americans, as well as to demonstrate the ways GIS-specific methodologies may present unique challenges, considerations, and opportunities through their embedded assumptions and epistemological peculiarity as data-driven Cartesian world-views.

	Chinese	Korean	Malaysian	Vietnamese
Foreign born	63%	62%	83%	64%
U.S. citizen†	58%	60%	24%	75%
Living in the U.S. more than 10 years†	66%	74%	48%	76%
High school or less	31%	25%	27%	48%
Bachelor's degree or more	53%	54%	60%	29%
Speak English proficiently	59%	63%	70%	51%
Median age (in years)	36	35	30	36
Married (among adults)	58%	56%	43%	55%
Women ages 15 to 44 who gave birth in past 12 months	5%	5%	***	6%
Living in multigenerational household	25%	20%	11%	32%
Median household income	$70,000	$60,000	***	$60,000
Homeownership rate	62%	47%	***	65%
Living in poverty	14.40%	12.80%	27.70%	14.30%
Unemployment rate	5.80%	5.60%	3.80%	5.70%

Source: Pew Research Center, 2017


AMERICAN COMMUNITY SURVEY
Ethnic groups
SES indicators
CENSUS TRACT GEOGRAPHY
Join
Find clustering
NEW CLUSTERHOOD BOUNDARIES
Analyze socioeconomic indicators
COMPARISON OF SOCIOECONOMIC STATUS BETWEEN ETHNIC CLUSTERS

Methodology

Informed by the theoretical and historical framework for our research, we present the following sets of research questions as the starting point for our study:

Does disaggregated data map the spatial distribution/clustering of distinct ethnic immigrant communities? What type of socioeconomic patterns do these hypothetical clusterhood units reveal?

It is important to note that we are not able to gather disaggregated data on the aforementioned social indicators, disaggregated by Asian ethnic groups. In summary, our methodology explores a novel way of approaching this data limitation by examining the spatial relationships between 2 sets of data: "clusterhoods" of Asian American ethnic groups, and social indicator data. All data is from the Census Bureau, American Community Survey 2012 to 2016. Geometry is sourced from the TIGER files and New York Borough Boundaries shapefile.

After defining and creating clusterhoods as our primary operational unit to conceptualize Asian American communities (loosely disaggregated by ethnicity) and using these units to summarize indicator data, we will briefly reflect on the geographies of a small subset of these clusterhoods to assess the analytical soundness of using the Getis-Ord Gi* cluster algorithm to create clusterhoods. The algorithm proves to be efficient for surface-level analysis of some communities on some scales, but due to its assumptions on sociocultural practices of Asian Americans, falls short in some analyses.

Indicator data

We gather tabular attribute data representing a limited set of estimated demographic data by all census tracts in the five boroughs of New York City: total population count, count of population over 25 without a high school diploma (we will refer to as "No HS Diploma"), high rent burden, extreme rent burden, and mean commute time to work (in minutes). In addition, we gather the tally of the total population of all ethnic groups. This tally information is used to create the geometry for our clusterhoods. With our indicator data joined to our clusterhoods for each Asian ethnic group, we begin our exploratory GIS.

Explorations

Our clusterhoods seem to confirm patterns of Asian American communities that we anticipated. For example, we know from common knowledge that areas of Southern Brooklyn and Northeast Queens contain numerous communities of Chinese Americans, and that Flushing-Bayside contains numerous Korean American communities.

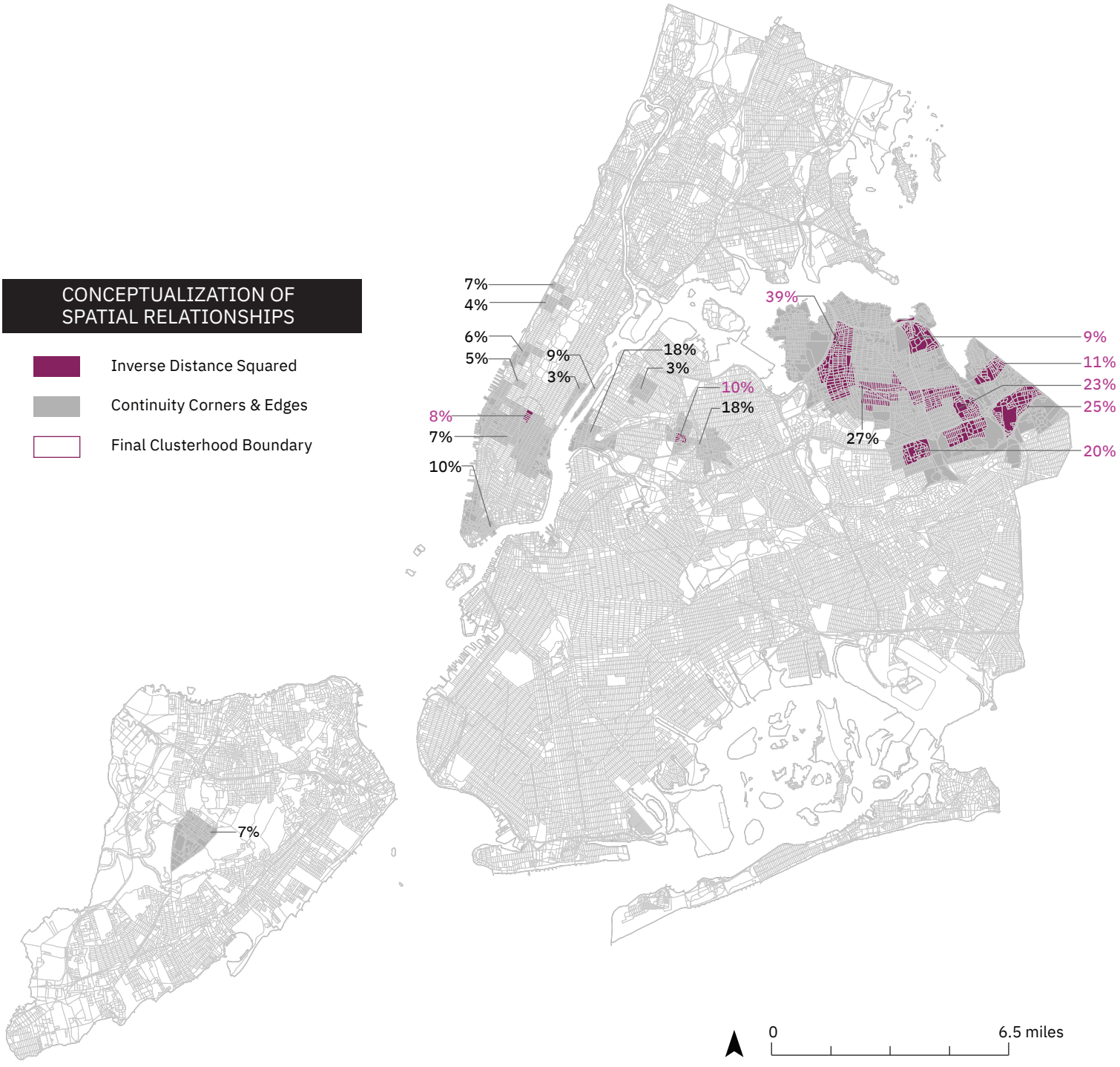
After identifying our clusterhoods, we looked at the socioeconomic statistics for each clusterhood. As we noted earlier, these socioeconomic statistics are not specific to each ethnic group - they reflect the socioeconomic statistics of the total population of the census tracts included within our clusterhood boundaries. In our analysis we include a percentage of the ethnic population over the total population of the census tracts.

We use education attainment (% of people with no high school diploma over the age of 25), high rent burden (% of people who spend more than 30% of their income on rent), extreme rent burden (% of people who spend more than 50% of their income on rent) and mean commute time to work as a measure of socioeconomic risk.

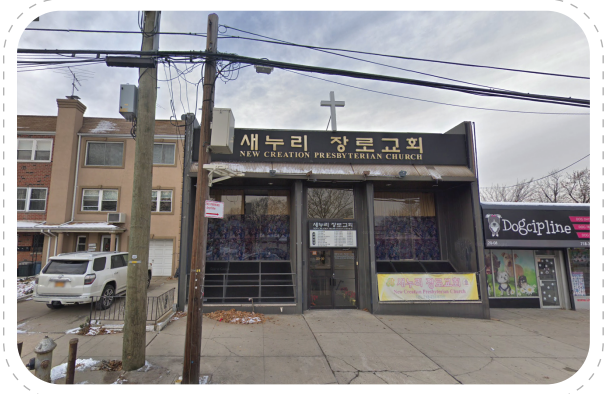
Upon examination of our results, we found that certain ethnic groups had higher percentages of socioeconomic risk than others (confirming our initial hypothesis that the "Asian" and model minority monolith obscures the struggles of certain ethnic groups). For the purposes of this report, we have isolated four ethnic groups: Chinese, Korean, Malaysian and Vietnamese and highlighted three clusterhoods for each group.

Conceptualization of Spatial Relationships

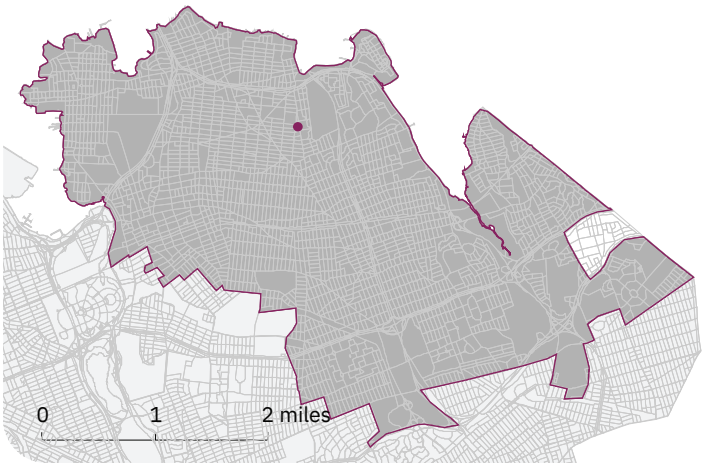
Using the Korean clusterhoods as an example, we illustrate our decision making process when choosing clusterhood boundaries. From ArcMap, we output two different "conceptualizations of spatial relationship" options from the Getis Ord Gi * clustering test. One was "Contiguity Edges and Corners" (CEC) and the other was "Inverse Distance Squared" (Inv Dist). CEC offers a more generous area - selecting all census tracts that are contiguous with the "hot spot" of the selected ethnic group cluster. The Inv Dist option has a faster falloff rate, and only selects the census tracts with a high population count of the selected ethnic group.



In selecting between these options, we made assumptions about the effect that smaller or larger clusters of ethnic populations could have on the built environment and social fabric of the neighborhood. We assumed that larger populations would have more of an effect on the built environment (selecting CEC) and that smaller populations would not (Inv Dist) although the social connections are still significant.



INVERSE DISTANCE SQUARED

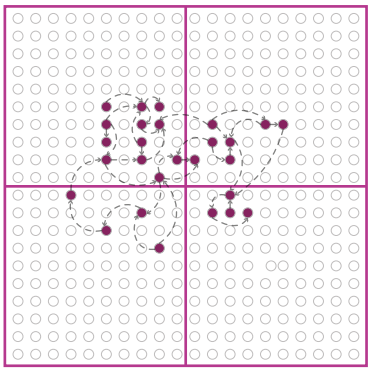


CONTIGUITY EDGES AND CORNERS

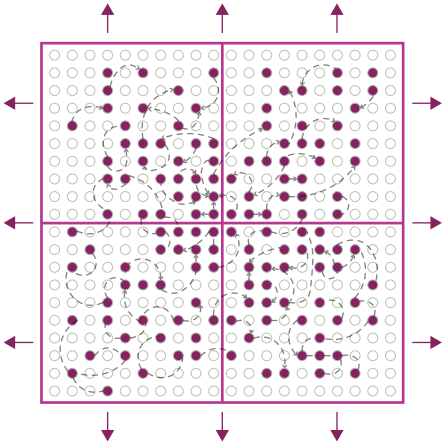
EFFECT OF POPULATION SIZE ON THE BUILT ENVIRONMENT

- In ethnic group
- Not in ethnic group
- Social connections
- Built environment effects
- Census tract boundary

SMALL ETHNIC GROUP POPULATION PER CENSUS TRACT

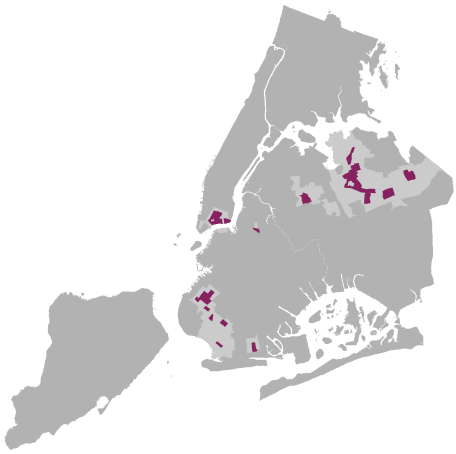


LARGE ETHNIC GROUP POPULATION PER CENSUS TRACT



Clusterhood Analysis: Chinese

Within the Chinese community, we see a polarization of education attainment. While the percentage of adults without a high school diploma is relatively low in Flushing and Chinatown (5% and 2%, respectively), 23% of adults in Bensonhurst do not have a highschool diploma.



Clusterhood 5 Flushing

Percent Chinese*	36%
Total population (Chinese)	119,484
Total population (Asian)	185,044
Total population	334,470

SOCIOECONOMIC INDICATORS ACROSS ALL POPULATIONS WITHIN CLUSTERHOOD

Age 25+ without HS diploma	5%
Rent burden (high)	10%
Rent burden (extreme)	6%
Mean commute to work time	39 minutes

Clusterhood 3 • Chinatown

Percent Chinese*	39%
Total population (Chinese)	43,525
Total population (Asian)	51,437
Total population	110,363

SOCIOECONOMIC INDICATORS ACROSS ALL POPULATIONS WITHIN CLUSTERHOOD

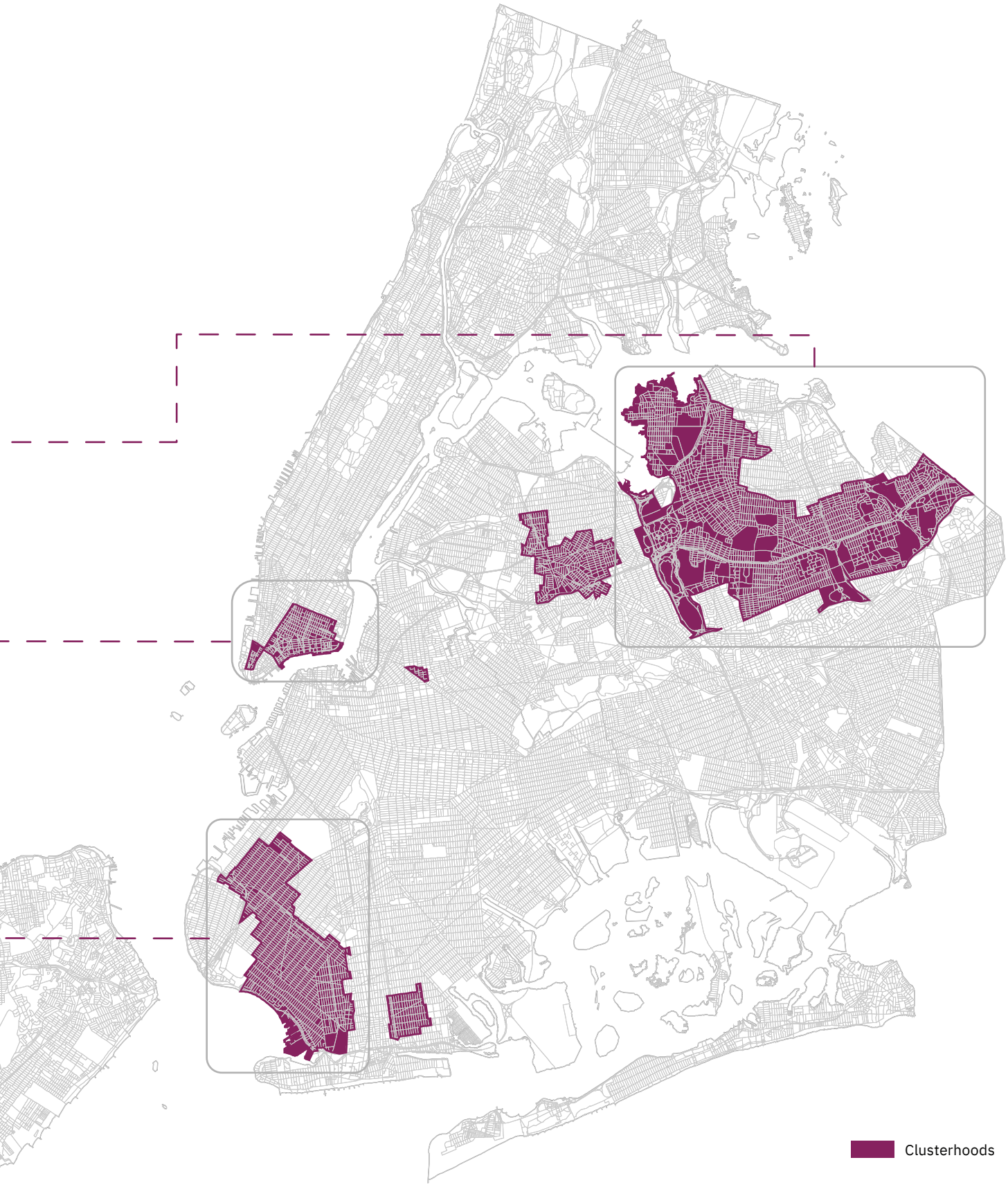
Age 25+ without HS diploma	2%
Rent burden (high)	18%
Rent burden (extreme)	9%
Mean commute to work time	30 minutes

Clusterhood 2 • Bensonhurst

Percent Chinese*	36%
Total population (Asian)	155,180
Total population (Chinese)	135,235
Total population	374,930

SOCIOECONOMIC INDICATORS ACROSS ALL POPULATIONS WITHIN CLUSTERHOOD

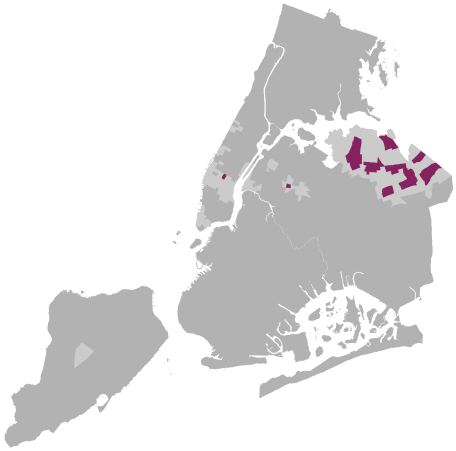
Age 25+ without HS diploma	23%
Rent burden (high)	12%
Rent burden (extreme)	7%
Mean commute to work time	45 minutes



* percent of total population in clusterhood who reported “Chinese.” Taiwanese are not included in the Census data under “Chinese.”

Clusterhood Analysis: Korean

Compared to other Asian ethnic groups, the Korean community in NYC has high education attainment (a low percentage of adults without a high school diploma: 1% in Koreatown, 10% in Woodside and 13% in Bayside/Flushing) and relatively low rent burden (24% in Koreatown, 16% in Woodside and 8% in Bayside/Flushing), indicating that they are a group pushing this image of the "model minority."



Clusterhood 3 • Koreatown

Percent Korean*	15%
Total population (Korean)	546
Total population (Asian)	973
Total population	2,065

SOCIOECONOMIC INDICATORS ACROSS ALL POPULATIONS WITHIN CLUSTERHOOD

Age 25+ without HS diploma	1%
Rent burden (high)	24%
Rent burden (extreme)	15%
Mean commute to work time	21 minutes

Clusterhood 2 • Woodside

Percent Korean*	10%
Total population (Korean)	713
Total population (Asian)	2,319
Total population	5,043

SOCIOECONOMIC INDICATORS ACROSS ALL POPULATIONS WITHIN CLUSTERHOOD

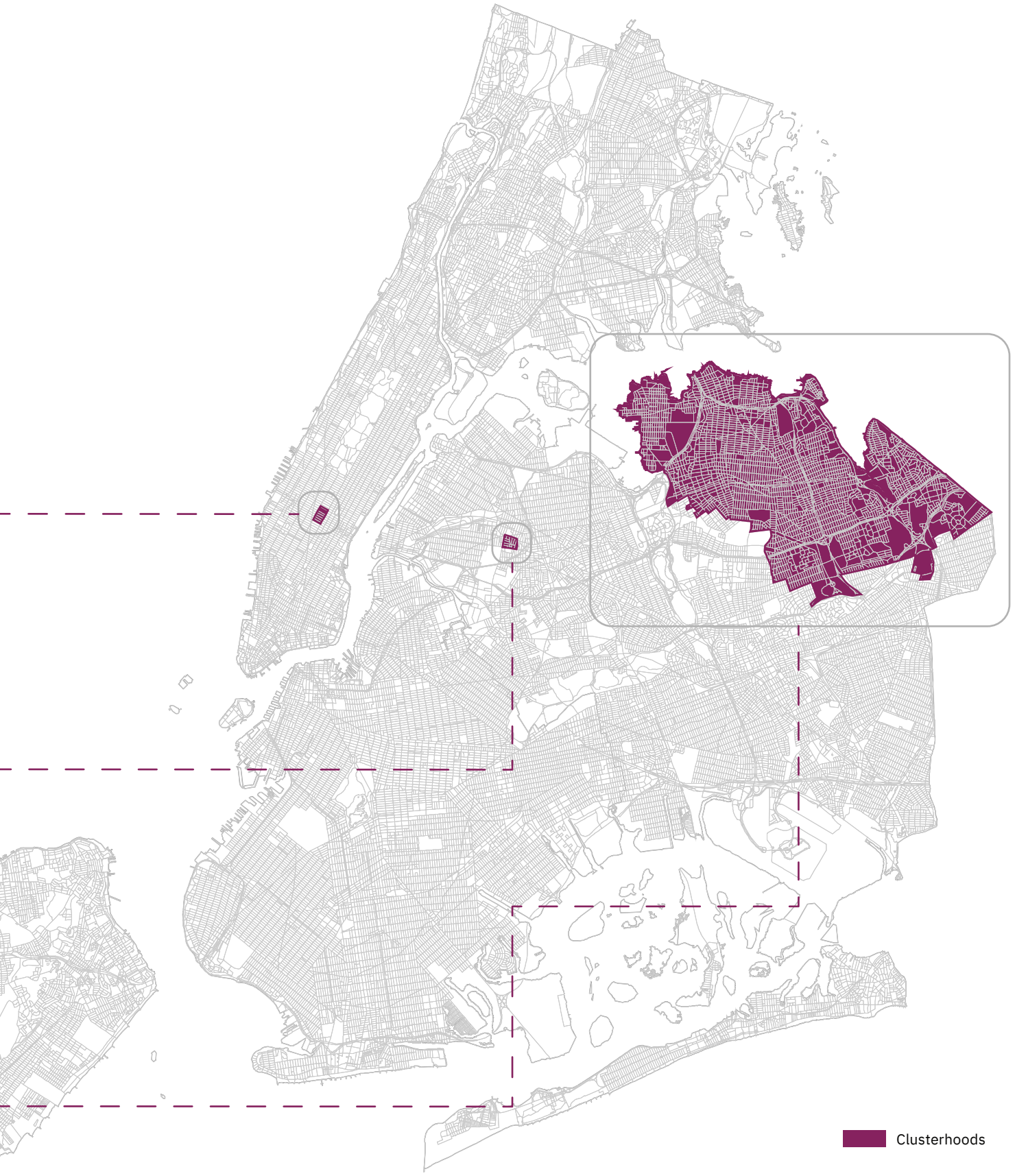
Age 25+ without HS diploma	10%
Rent burden (high)	16%
Rent burden (extreme)	9%
Mean commute to work time	37 minutes

Clusterhood 1 • Bayside and Flushing

Percent Korean*	11.7%
Total population (Korean)	41,258
Total population (Asian)	169,889
Total population	352,337

SOCIOECONOMIC INDICATORS ACROSS ALL POPULATIONS WITHIN CLUSTERHOOD

Age 25+ without HS diploma	13%
Rent burden (high)	8%
Rent burden (extreme)	5%
Mean commute to work time	40 minutes



* percent of total population in clusterhood who reported “Chinese”

Clusterhood Analysis: Malaysian

Similar to the Chinese community, we see a polarization of education attainment. Sunset Park has the highest percentage of adults without a highschool diploma (41% in Sunset Park compared to 22% in Flushing and Chinatown) among the clusterhoods we highlighted across all ethnicities.

Clusterhood 5 • Flushing/Queensboro Hill

Percent Malaysian*	6%
Total population (Malaysian)	308
Total population (Asian)	40,725
Total population	53,918

SOCIOECONOMIC INDICATORS ACROSS ALL POPULATIONS WITHIN CLUSTERHOOD

Age 25+ without HS diploma	22%
Rent burden (high)	15%
Rent burden (extreme)	9%
Mean commute to work time	41 minutes

Clusterhood 3 • Chinatown

Percent Malaysian*	6%
Total population (Malaysian)	433
Total population (Asian)	38,181
Total population	74,103

SOCIOECONOMIC INDICATORS ACROSS ALL POPULATIONS WITHIN CLUSTERHOOD

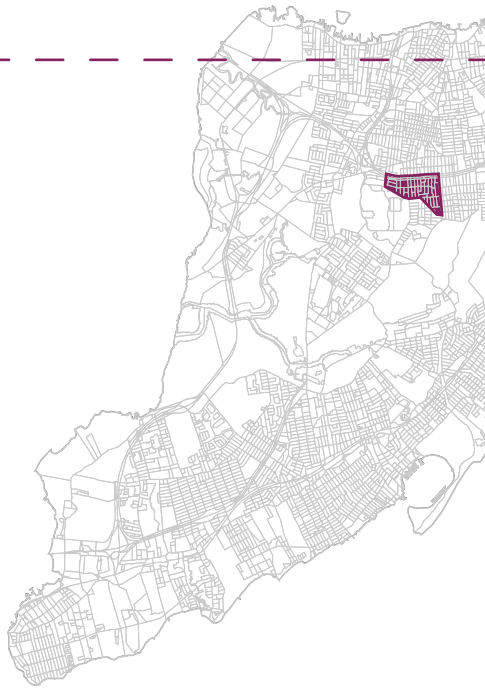
Age 25+ without HS diploma	22%
Rent burden (high)	18%
Rent burden (extreme)	9%
Mean commute to work time	30 minutes

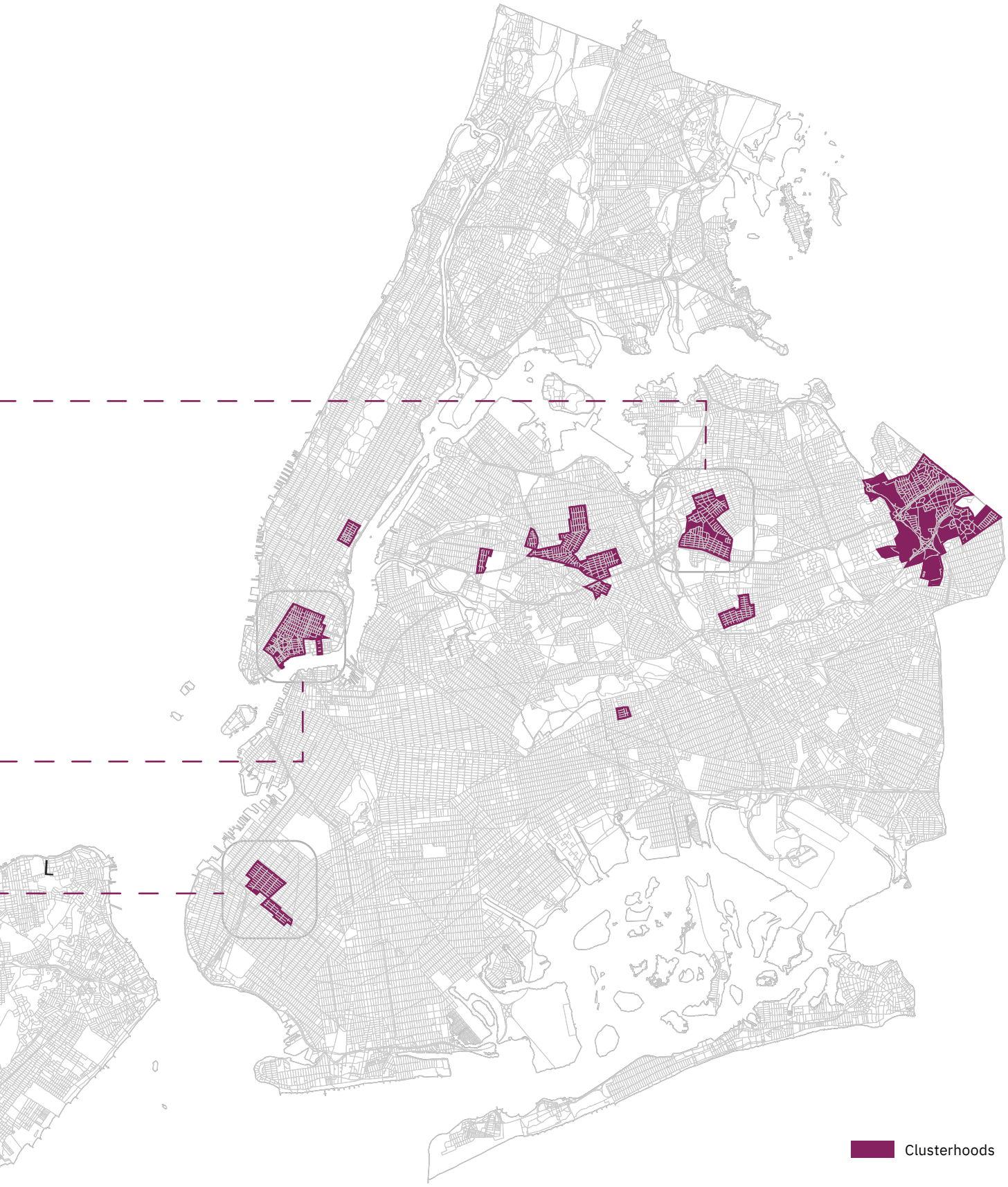
Clusterhood 2 • Sunset Park

Percent Malaysian*	0.8%
Total population (Malaysian)	247
Total population (Asian)	21,171
Total population	29,338

SOCIOECONOMIC INDICATORS ACROSS ALL POPULATIONS WITHIN CLUSTERHOOD

Age 25+ without HS diploma	41%
Rent burden (high)	13%
Rent burden (extreme)	7%
Mean commute to work time	45 minutes

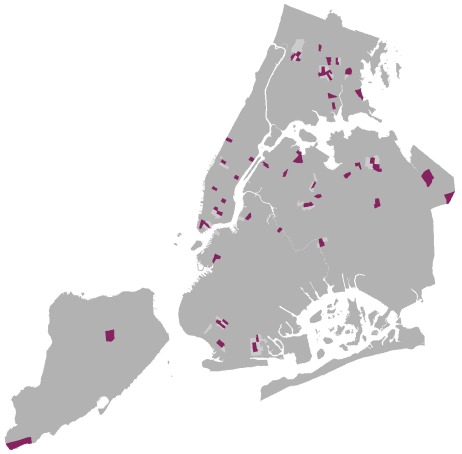




* percent of total population in clusterhood who reported “Malaysian”

Clusterhood Analysis: Vietnamese

The Vietnamese clusterhoods were the most dispersed out of the ethnic groups we selected. For example, out of all our selected ethnic groups, the Vietnamese was the only group that had a clusterhood in the Bronx. The socioeconomic statistics remain relatively consistent throughout the clusters we selected (with a slightly lower rent burden in the Bronx).



Clusterhood 17 • Bronx

Percent Vietnamese*	4.2%
Total population (Vietnamese)	701
Total population (Asian)	2,322
Total population	16,624

SOCIOECONOMIC INDICATORS ACROSS ALL POPULATIONS WITHIN CLUSTERHOOD

Age 25+ without HS diploma	14%
Rent burden (high)	13%
Rent burden (extreme)	8%
Mean commute to work time	40 minutes

Clusterhood 24 • Flushing

Percent Vietnamese*	2.6%
Total population (Vietnamese)	210
Total population (Asian)	6,008
Total population	8,073

SOCIOECONOMIC INDICATORS ACROSS ALL POPULATIONS WITHIN CLUSTERHOOD

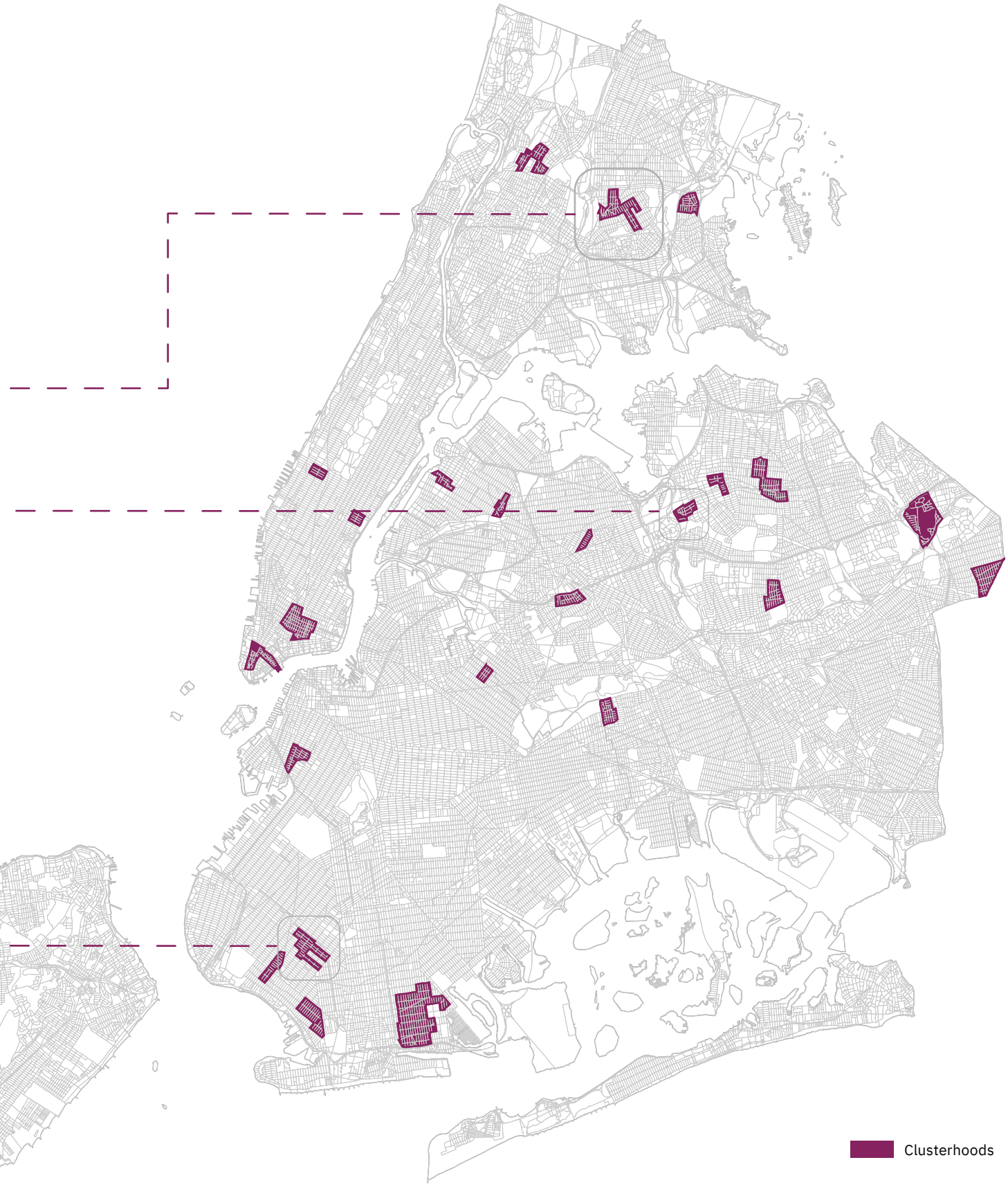
Age 25+ without HS diploma	23%
Rent burden (high)	19%
Rent burden (extreme)	15%
Mean commute to work time	40 minutes

Clusterhood 22 • Bensonhurst

Percent Vietnamese*	1.4%
Total population (Vietnamese)	112
Total population (Asian)	7,922
Total population	15,535

SOCIOECONOMIC INDICATORS ACROSS ALL POPULATIONS WITHIN CLUSTERHOOD

Age 25+ without HS diploma	23%
Rent burden (high)	11%
Rent burden (extreme)	6%
Mean commute to work time	46 minutes



* percent of total population in clusterhood who reported “Vietnamese”

Bibliography

Aina, Sefa. "The Significance of Data Disaggregation to the Asian American and Pacific Islander (AAPI) Community." Whitehouse.Gov, 27 July 2012, https:// obamawhitehouse.archives.gov/blog/2012/07/27/significance-data-disaggregation-asian-american-and-pacific-islander-aapi-community.

Dizon, Jude Paul Matias. "Lessons on ethnic data disaggregation from the "count me in" campaign." The Vermont Connection 32.1 (2011): 3.

NYU Center for the Study of Asian American Health. Cambodian Community Health Resources and Needs Assessment (CHRNA). 2016, https://med.nyu.edu/asianhealth/research/community-health-resources-needs-assessment/2013-2016-community-reports.

Omi, Michael, and Howard Winant. Racial formation in the United States. Routledge, 2014.

Trinh-Shevrin, Chau, et al. "Defining an integrative approach for health promotion and disease prevention: a population health equity framework." Journal of health care for the poor and underserved 26.2 0 (2015): 146. "NY State Assembly Bill A7352." NY State Senate, 24 Apr. 2017, https://www.nysenate. gov/legislation/bills/2017/a7352/amendment/original.

Data Sources

New York City Department of City Planning, *Neighborhood Tabulation Areas. [dataset].* 2018. <https://www1.nyc.gov/site/planning/data-maps/open-data/dwn-nynta.page>

New York City Department of City Planning. *New York City Borough Boundaries [shapefile].* Bytes of the Big Apple, Issue 16D. 27 October 2016. <http://www1.nyc.gov/site/planning/datamaps/open-data/districts-downloadmetadata.page>

US Census Bureau. *Asian alone or in any combination by selected groups (Table B02018)* , 2016 American Community Survey 1-Year Estimates. Retrieved via Census API using censusapi R package.

US Census Bureau. *Educational attainment (Table S1501), 2012-2016 American Community Survey 5-year estimates.* Retrieved via Census API using censusapi R package.

US Census Bureau. *Gross rent as a percentage of household income in the past 12 months (Table B25070), 2012-2016 American Community Survey 5-year estimates.* Retrieved via Census API using censusapi R package.

US Census Bureau. *Selected economic characteristics (Table DP03), 2012-2016 American Community Survey 5-year estimates. Retrieved via Census API using censusapi R package.*

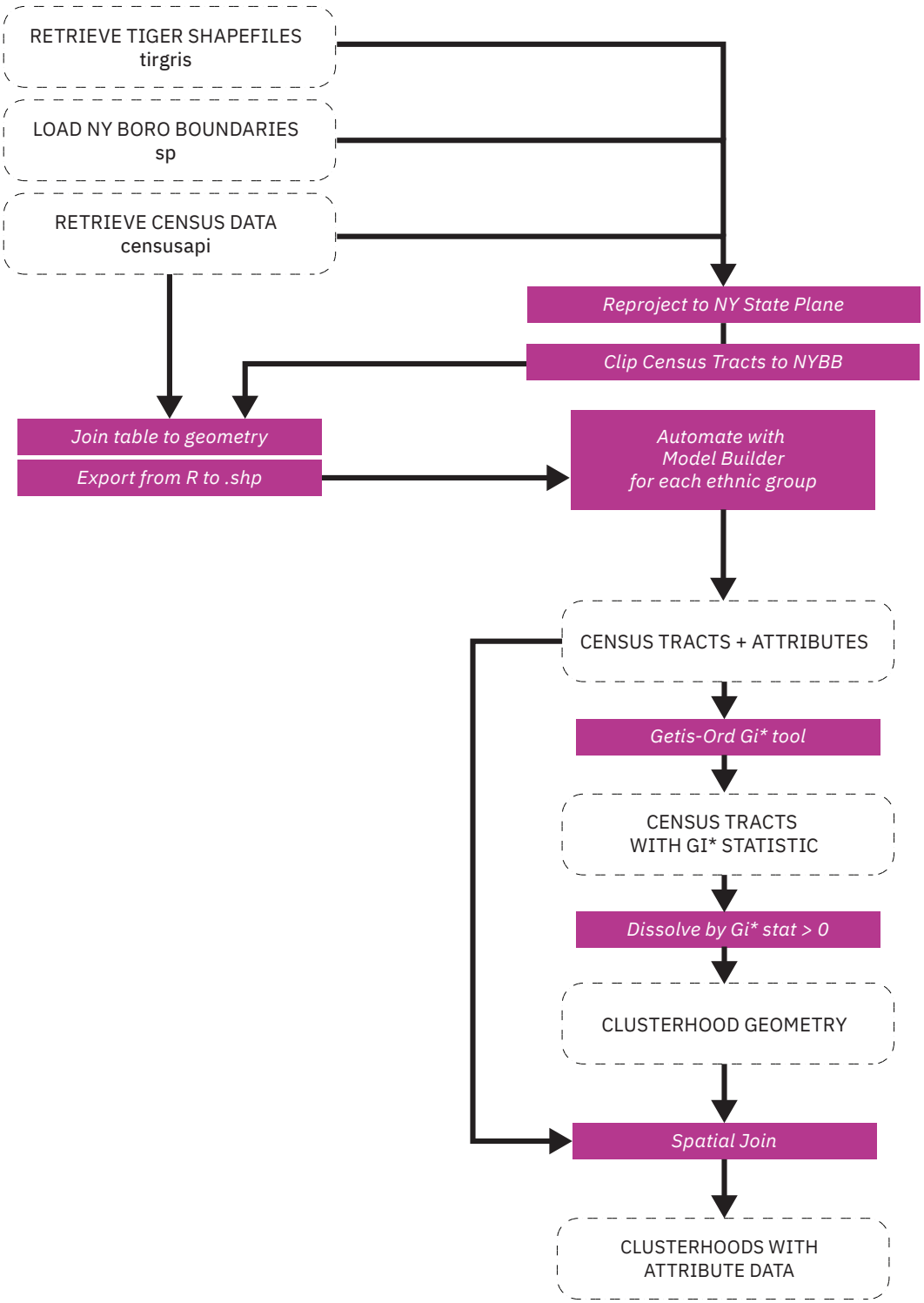
US Census Bureau. *TIGER/Line Shapefiles and TIGER/Line Files.* 2018. Retrieved via tigris R package.

Appendix: Workflow

Our process involves uses an extensive suite of tools that allow mostly automated, replicable results. With the exception of ArcGIS's Getis-Ord Gi* script, all tools used for gathering and joining data are free and open source. First, all Census data is collected via the official Census Bureau API using the censusapi package for R. R is used for both field calculations and renames, as well as rudimentary spatial operations such as clips and joins. Using R, appropriate fields are renamed and calculated from the downloaded Census tables. TIGER files are conveniently retrieved via tigris package for R, clipped to the New York Borough boundaries using the raster R package, and joined to prepared Census data via the sp package. Before census tract geometry is clipped to borough boundaries, the feature classes are reprojected using the sp package to the New York State Plane Long Island FIPS US Feet coordinate reference system (CRS) known in the spatialreference.org database as ESRI:102718. This file is output to a shapefile and OGR GeoPackage file for compatibility with QGIS and other GIS suites. Regardless of format, the primary output from this process is a single feature class containing geometry of census tracts clipped to borough boundaries, associated with an attribute table describing population of Asian ethnic groups as well as social indicator data.

After all shapefile data was imported into ArcMap, we applied the Getis-Ord Gi* tool to the census tract data, creating outputs using all 19 Asian ethnic groups as input fields. This process identifies statistically significant clustering that we use to define clusterhoods. Most notably, this process involves the key choice of Conceptualization of Spatial Relationships. The two we chose to explore were Nearest Neighbor (NN) and Contiguity Edges and Corners (CEC). After all cluster tests are complete, we result in 19*2 feature classes representing significant clusters. In order to operate on this geometry; for each feature class representing a single ethnic group, we select only statistically significant census tracts, and dissolve these into contiguous shapes that we call clusterhoods.

Unfortunately, the cluster tests and dissolve operations destroy our indicator data previously present in the input feature class, requiring us to spatial join our indicator data back to our new clusterhood geometry. This would consist of a total of 19 * 2 extremely tedious spatial join operations. To avoid this, we used ArcGIS ModelBuilder tool to automate the spatial join operations to bring our indicator data back to our clusterhood geometry.


RETRIEVE TIGER SHAPEFILES
tirgris
LOAD NY BORO BOUNDARIES
sp
RETRIEVE CENSUS DATA
censusapi
Reproject to NY State Plane
Clip Census Tracts to NYBB
Join table to geometry
Export from R to .shp
Automate with
Model Builder
for each ethnic group
CENSUS TRACTS + ATTRIBUTES
Getis-Ord Gi* tool
CENSUS TRACTS
WITH GI* STATISTIC
Dissolve by Gi* stat > 0
CLUSTERHOOD GEOMETRY
Spatial Join
CLUSTERHOODS WITH
ATTRIBUTE DATA